

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### Sufficient conditions for emergent synchronization in protocell models

Carletti, Timoteo; Serra, Roberto; Poli, Irene; Villani, Marco; Filisetti, Alessandro

*Published in:*

Journal of Theoretical Biology

*Publication date:*

2008

*Document Version*

Early version, also known as pre-print

[Link to publication](#)

*Citation for pulished version (HARVARD):*

Carletti, T, Serra, R, Poli, I, Villani, M & Filisetti, A 2008, 'Sufficient conditions for emergent synchronization in protocell models', *Journal of Theoretical Biology*, vol. 254, no. 4, pp. 741-751.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# SUFFICIENT CONDITIONS FOR EMERGENT SYNCHRONIZATION IN PROTOCELL MODELS

T. CARLETTI, R. SERRA, I. POLI, M. VILLANI AND A. FILISETTI

**ABSTRACT.** In this paper we study general protocell models aiming to understand the synchronization phenomenon of genetic material and container productions, a necessary condition to ensure sustainable growth in protocells and eventually leading to Darwinian evolution when applied to a population of protocells.

Synchronization has been proved to be an emergent property in many relevant protocell models in the class of the so-called Surface Reaction Models, assuming both linear and nonlinear dynamics for the involved chemical reactions. We here extend this analysis by introducing and studying a new class of models where the relevant chemical reactions are assumed to occur inside the protocell, in contrast with the former model where the reaction site was the external surface.

While in our previous studies the replicators were assumed to compete for resources, without any direct interaction among them, we here improve both models by allowing linear interaction between replicators: catalysis and/or inhibition. Extending some techniques previously introduced, we are able to give a quite general analytical answer about the synchronization phenomenon in this more general context. We also report on results of numerical simulations to support the theory, where applicable, and allow the investigation of cases which are not amenable to analytical calculations.

## 1. INTRODUCTION

Several attempts are currently under way to obtain protocells capable of growth and duplication, endowed with some limited form of genetics (Oberholzer et al., 1995; Rasmussen et al., 2004; Szostak et al., 2001; Mansy et al., 2008). The interest for these systems is motivated either by the quest to understand which are the minimal requirements for a life form to exist and evolve, or by the search for indications about the way in which primitive life might have emerged on Earth.

In order to study how protocells can develop, given that they do not yet exist, it is necessary to consider “simplified models able to capture universal behaviours, without carefully adding complicating details” (Kaneko, 2006). A protocell should comprise at least one kind of “container” molecule (typically a lipid or amphiphile) and one kind of replicator molecule – loosely speaking “genetic material”, hereafter called, Genetic Memory Molecule, GMM for short. This is typically a linear polymer which can be copied or a system of two or more kinds of replicators which catalyze each other’s synthesis – e.g. proteins and nucleic acids. There are therefore two kinds of reactions which are crucial for the working of the protocell, which in this paper will be called “key” reactions: those which synthesize the container molecules and those which synthesize the GMM replicators.

The two key reactions may take place at different rates. However, to achieve sustained protocell growth and avoid death by dilution, it is necessary that the two

proceed at equal rates, i.e. that the genetic material has doubled when the protocell splits into two – a condition referred to as *synchronization*, to ensure that each offspring will contain the same amount of genetic material as the mother. Indeed, if replication were slower than duplication, the concentration of genetic material would eventually vanish (we refer to the splitting of a protocell as duplication, and to the doubling of genetic polymers as replication). In the opposite case its concentration would grow unbounded. Of course, the requirement of doubling of the genetic material at duplication time refers to the average behaviour, while each single event is affected by noise and random fluctuations.

Note that synchronization has a further important property, namely that, even in the case where the kinetic equations for the GMMs have sub linear growth terms (Rasmussen et al., 2003; Munteanu et al., 2006), it leads to exponential growth of the population of protocells (a straightforward consequence of constant doubling time <sup>1</sup>) and therefore to strictly Darwinian selection among protocells.

Most of the different protocell architectures which have been proposed can be divided in two main families, according to the region of cell space where these key reactions occur. Some proposals, which have been called *Surface Reaction Models* (Serra et al., 2007a) – SRM for short, assume that the key reactions take place on the surface of the cell membrane, as hypothesized for the Los Alamos bug (Rasmussen et al., 2003, 2004). Other architectures exist, for instance the RNA-cell (Oberholzer et al., 1995; Szostak et al., 2001), where the key reactions develop in the interior of the vesicle. For this reason we call our model, whose inspiration has been drawn from this latter case, *Internal Reaction Model* – IRM for short.

In this paper we address the synchronization question for both proposed architectures, exhibiting a unified analysis; in fact we are able to prove that working with quantities (of chemicals) instead of concentrations allows us to map one model on the other and thus to provide a unified view.

In the case of SRM, we are also able to consider cases (see § 3.4) where the “genetic molecules” are actually the same lipids that compose the “container”, allowing us to consider models close to the GARD – model (Segré et al., 1998) or to the one proposed by Kaneko and Yomo (2002) although in this paper we limit ourselves to considering only linear interactions. In these models the (compositional) information is carried by the diversity of lipids in the vesicle or micelle, thus the synchronization problem here can be restated in terms of the reproduction of the whole set of molecules before division occurs, so as to guarantee the maintenance of information content.

The problem of synchronization has already been studied in previous works by means of a class of abstract surface reaction models of protocells (Serra et al., 2007a,b) and it has been shown that in several cases synchronization is an emergent property, in the sense that, through successive generations of protocells, the doubling times of both container and replicators, tend asymptotically to the same value even if at the beginning they were different. This was contrasted to earlier models, like the well-known Chemoton (Carletti and Fanelli, 2007; Gánti, 1997; Munteanu and Solé, 2006), where synchronization was achieved by *ad hoc* hypotheses concerning the form of the kinetic equations.

---

<sup>1</sup>Here we ignore further terms which might limit the growth of the whole population of protocells, e.g. competition for limited resources or growth in a limited volume.

In models involving a single Genetic Memory Molecule, synchronization is always achieved once the growth of the lipid container is linear with respect to the quantity of the replicator (Serra et al., 2007a). This result has been generalized to models where the replicator equation is non-linear or when the growth of the container is given by a non-linear function of the amount of genetic material (Serra et al., 2007b).

In models where more than one GMM coexist in the same protocell, but limiting the treatment to the case where there is no direct interaction among them, synchronization was achieved: if the replicator kinetics is linear, only the fastest replicator asymptotically survives, while if it is parabolic there is coexistence of different replicators in the long time limit (Serra et al., 2007a).

The fact that several different hypotheses lead asymptotically to synchronization raises the question whether this is a general property of the SRM or even of a larger class including also the IRM. In the present paper we therefore explore this wider class of models taking into account direct interaction, positive and negative, among the replicators. We consider the case of linear replication kinetics, finding sufficient conditions to guarantee synchronization: note however that, since protocell division is taken into account, the overall model is non-linear, so its analysis is far from being trivial. We are aware that such assumptions limit the application of our method to a limited class of models, and that relaxing them we cannot obtain such analytical results, nevertheless we support our choice with the following two reasons. First, we have proved (Serra et al., 2007b) that synchronization arises under general assumptions of non-linear coupling between container growth and GMMs or non-linear kinetics for GMMs replication; second, we stress once again that we are looking for simplified models able to capture universal features, neglecting specific, model-dependent details, hence the linear assumptions are a reasonable starting point. For this same reason we here neglect higher order phenomena like diffusion and permeation processes, whose influence can be important but whose analyses go beyond the scope of the present work.

The treatment of the subject is mostly analytical; we nevertheless present some numerical simulations both to support our results and to explore cases where the rigorous analysis cannot be performed.

The paper is organized as follows. In section 2 we will briefly introduce the two protocell architectures that somehow inspired our models: the Los Alamos bug and the RNA-cell. Then we will introduce our models: the surface reaction model and the internal reaction model; and finally we will discuss the relevant equations describing their dynamics. Section 3 will contain a full analysis of the dynamics of these models and the proof that synchronization can be achieved, provided some conditions on the involved coefficients are satisfied. Finally in section 4 an in-depth discussion of these conditions and of their physical meaning will be provided together with some comments on possible further directions of research.

## 2. TWO PROTOCELLS MODELS

The aim of this section is to introduce our models describing two possible architectures for living protocells, inspired by some current bio-chemical researches. First for the sake of completeness we will briefly introduce the models and then our approach will follow.

**2.1. Two possible artificial minimal cells.** According to Rasmussen et al. (Rasmussen et al., 2004) the Los Alamos Bug is a synthetic organism that integrates three functionalities: a lipid container, a photo-metabolic system and a hydrophobically anchored templating polymer that influences metabolic kinetics.

The role of the proto-container is to hold together the other two key aggregates. Moreover, authors assumed that the interior lipid phase as well as the water/lipid interface possess very different physicochemical properties with respect to bulk water, in such a way that the high concentrations determined by the spatial proximity of anchored molecules will enhance the chemical reactions: i.e., both the lipid phase and the lipid/water interface act as catalysts.

The assumed GMMs are lipophilic PNA or PNA-like nucleic acids; this choice has been motivated by the stronger interactions with the lipid phase of such molecules, thanks to their hydrophobic backbone; moreover PNA is more plausible than RNA or DNA in terms of prebiotic synthesis (Nelson et al., 2000). These polymers, in the single-stranded template form, are located at the lipid/water surface, exposing the hydrophilic bases to the aqueous medium while the hydrophobic backbone sinks into the lipid layer. The system is fed with oligos from the aqueous phase so that double-stranded templates can be formed – a reaction that is assumed to be driven by visible light as the primary energy source – and during this process new lipids are also produced. Such lipids will be immediately incorporated in the container triggering its growth, while the double-stranded templates will eventually dissociate into its two strands, still remaining anchored to the lipid/water interface. Moreover, the dynamical characteristics of the reactions are determined by the PNA bases sequences, hence their role as genetic memory molecules.

A somehow different approach has been proposed in (Oberholzer et al., 1995; Szostak et al., 2001; Mansy et al., 2008), where the starting point is the design and construction in the laboratory of a RNA replicase – an RNA molecule that can act both as a template for the storage and transmission of genetic information, and as an RNA polymerase that can replicate its own sequence (Hager et al., 1996; Bartel, 1999; Bartel and Unrau, 1999). This complex must then be incorporated into some form of compartment, to increase the replication rate but also because in this way advantageous mutations can lead to preferential replication: after replications, mutations and random assortment, some compartments will be occupied by mutant replicases that can replicate each other more efficiently, giving them an overall advantage (evolution). Authors thus hypothesized the existence of a lipid vesicle membrane, able to self-assemble thanks to the interactions between the available lipid molecules.

The last step to obtain a living organism is to introduce some advantage in survival, growth or replication for the membrane component, directly related to the RNA molecule, for instance a ribozyme that synthesizes amphipathic lipids pushing the membrane to grow. Once the container and the genetic material are coupled the whole protocell will evolve as a whole, and improved ribozymes – because of favorable mutations – would have a growth and replication advantage.

**2.2. Surface reaction models of protocells.** For the sake of completeness in a first part of this section we recall the model and the main results concerning the surface reaction models in the case where only one kind of genetic molecule is present; we refer the interested reader to (Serra et al., 2007a) for more details and

demonstrations. Then we generalize these models to the case of  $N$  kinds of GMMs interacting with each other.

So let us first consider the case where there is a single replicator, represented by some generic  $X$ -molecule in the protocell lipid phase <sup>2</sup>, and let its quantity (i.e. number of moles) be denoted by  $X$ . Let also  $C$  be the total quantity of “container” (i.e., lipid membrane forming vesicles or micelles) and  $V$  its volume, which is equal to  $C/\rho$  (where  $\rho$  is the density, assumed to be constant). Finally, let  $S$  denote the surface area, which is a function of  $V$  ( $S$  is approximately proportional to  $V$  for a large vesicle with a very thin bilayer membrane, a condition that will be referred to as the “thin vesicle case”, and to  $V^{2/3}$  for a micelle).

We assume, according to the Labug hypothesis, that the  $X$ -molecule favors the formation of amphiphiles, and that only the fraction which is near the external surface is effective in doing so, since precursors are found outside the protocell. We also assume that the replication of the  $X$ -molecule takes place near the external surface. For example, if the latter is a self-replicating linear polymer, its replication is accomplished by synthesizing a complementary chain from free activated monomers. However, we will develop a treatment which allows greater generality than pure self-replication.

Following the standard assumptions already used and discussed in (Serra et al., 2007a,b), namely:

- (1) spontaneous amphiphile formation is negligible, so that only the catalyzed term matters;
- (2) the precursors (both of amphiphiles and templates) are buffered;
- (3) the surface,  $S$ , is proportional to some power of the volume,  $V^\beta$  ( $\beta$  ranging between  $2/3$ , for a micelle, and  $1$ , for a very thin vesicle), and therefore also to the amount of container,  $C^\beta$ ;
- (4) diffusion is very fast within the lipid phase, so concentrations can be assumed to be homogeneous;
- (5) the protocell breaks into two identical daughter units when its container size reaches a certain threshold and then halving the genetic and container molecules between them;
- (6) the rate limiting step which may appear in the replicator kinetic equations does not play a significant role when the protocell is smaller than the division threshold;

one obtains the following approximate equations which describe the growth of a protocell between two successive divisions:

$$(2.1) \quad \frac{dX}{dt} = \eta C^{\beta-1} X \quad \text{and} \quad \frac{dC}{dt} = \alpha C^{\beta-1} X ,$$

where  $\eta$  and  $\alpha$  are two positive constants, denoting respectively the rate of self replication of genetic molecules and the container growth. When  $C$  reaches a critical threshold, the cell breaks into two equal daughter protocells and each one will start the next division cycle with an initial amount of the  $X$ -molecule equal to one half of the value which it has attained at the end of the previous cycle (perfect halving hypothesis).

---

<sup>2</sup>This model is invariant with respect to the way in which either amount of  $C$ -molecules or  $X$ -molecules are measured; for example, if they were measured as number of molecules the equations would retain exactly the same form (of course, the units of the kinetic constants would be different).

The generalization to the case where there are more replicators is straightforward. Let:

$$(2.2) \quad \vec{X} = (X^{(1)}, X^{(2)}, \dots, X^{(N)}) ,$$

denote the total quantity (moles) of  $N$  different types of replicating molecules in the protocell lipid phase, called for short  $\vec{X}$ -molecules. Obviously, all the quantities  $X^{(i)}$  must be real and non negative for  $i = 1, \dots, N$ . The  $N$ -dimensional generalization of Eqs. (2.1) is then

$$(2.3) \quad \begin{cases} \frac{d\vec{X}}{dt} &= C^{\beta-1} M \vec{X} \\ \frac{dC}{dt} &= C^{\beta-1} \vec{\alpha} \cdot \vec{X} , \end{cases}$$

where  $\alpha = (\alpha_1, \dots, \alpha_N)$  is a vector with positive entries denoting the coupling term between the container growth and each replicator, while the (constant and real) matrix element  $M_{ij}$  represents the contribution of the  $X^{(j)}$ -molecule to the growth of the  $X^{(i)}$ -molecule. We also assume the matrix  $M$  to be invertible, to avoid redundancy of chemical reactions, and thus to limit the analysis to the minimal number of independent chemical species.

An important simplification can now be considered: as it was demonstrated in (Serra et al., 2007a), in order to determine whether there is synchronization in the asymptotic time limit, one can limit oneself to consider the  $\beta = 1$  case. The final result does not depend on  $\beta$ , while of course this parameter affects the speed with which it is approached; this essentially is a non-linear rescaling of time, useful to simplify the analysis. With this simplification, the basic equations (which are valid between two successive divisions) are then

$$(2.4) \quad \begin{cases} \frac{d\vec{X}}{dt} &= M \vec{X} \\ \frac{dC}{dt} &= \vec{\alpha} \cdot \vec{X} . \end{cases}$$

As outlined above, we assume that division takes place when the mass (or equivalently the volume, since density is assumed to be constant) of the protocell reaches a certain critical size. Without loss of generality we may then assume that the initial size of the protocell is one half of the final value (indeed, if the size of the very first protocell were different then it would suffice to consider the evolution from the following generation).

Let us observe here that this assumption, and also the halving hypothesis of genetic material at division, are not so restrictive as it could be perceived at first reading. In fact there is nothing magical about the chosen factor 2, any other factor strictly larger than one, would give the same results, except of course modifying the asymptotic values of the amount of  $\vec{X}$ -molecules and division time. As for the second assumption, we are aware that in the physical case, membrane splitting and divisions of  $\vec{X}$ -molecules, could be better described by introducing some probability distribution functions of divisions events, with mean value 1/2; however, this will not change our results except that the asymptotic values for  $\vec{X}_\infty$  and  $\Delta T_\infty$  will refer to averaged values of some distribution.

So, starting with an initial quantity of container  $C(T_0) = \theta/2$  at time  $T_0$ , we assume that once the container size  $C(t)$  reaches the critical value  $\theta$  it will divide into two equal protocells of size  $\theta/2$ . Let  $\Delta T_0$  be the time interval needed to double  $C$  from this initial condition, and let  $T_1 = T_0 + \Delta T_0$  be the time at which the critical mass  $\theta$  is reached. Clearly  $\Delta T_0$  is a function of the initial quantity of



replicators,  $\vec{X}_0$ . Let us denote  $\vec{X}(t; t_0, \vec{X}_0)$  the quantity of  $\vec{X}$ -molecules at time  $t$ , i.e. the solution of (2.4) with initial datum  $\vec{X}(T_0) = \vec{X}_0$ . When there will be no ambiguity with respect to the initial datum we are considering, we will use the shorter notation  $\vec{X}(t)$ . The final value of  $\vec{X}$  just before the division is thus  $\vec{X}(T_1; t_0, \vec{X}_0) \equiv \vec{X}(T_1)$ . By assumption each offspring will start the next cycle with an initial concentration of replicators equal to half the quantity present at the end of the previous cycle, i.e. in formula  $\vec{X}_1 = \vec{X}(T_1)/2$ , from this point till the next division the dynamics is determined again by (2.4), let us however observe that the solution could be different because the initial data have been set differently. Let us denote the successive doubling time by  $T_2 = T_1 + \Delta T_1$ , thus the third generation will start with an initial value of genetic material,  $\vec{X}_2 = \vec{X}(T_2)/2$ , and so on (see Fig. 1 for a cartoon describing the construction).

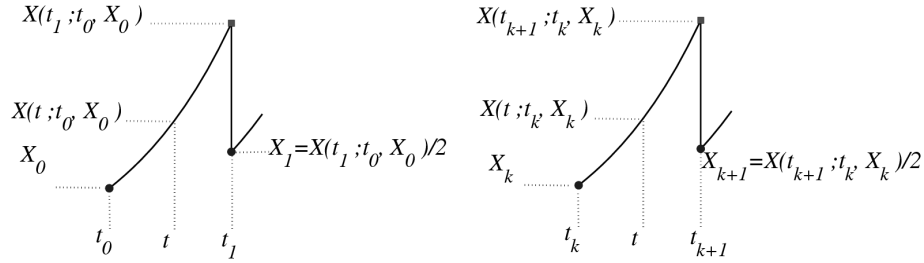


FIGURE 1. A schematic representation of the construction of the sequence  $\vec{X}_k$ .  $X(t; t_k, X_k)$  denotes the amount of  $X$ -molecule at time  $t$  during the continuous growth, starting from an amount  $X_k$  at time  $t_k$ . The division occurs at time  $t_{k+1}$  and the next protocell cycle will start with an amount of  $X$ -molecule given by  $X_{k+1} = X(t_{k+1}; t_k, X_k)/2$ .

We generalize the previous discussion with the following equations, which refer to the  $(k+1)$ -th cell division cycle that starts at time  $T_k$  and ends at time  $T_{k+1}$ :

$$(2.5) \quad \frac{\theta}{2} = \int_{T_k}^{T_{k+1}} \frac{dC}{dt} dt \quad \text{and} \quad \vec{X}_{k+1} = \frac{1}{2} \vec{X}(T_{k+1}) = \frac{1}{2} \vec{X}(T_{k+1}; T_k, \vec{X}_k),$$

provided  $T_{k+1}$  does exist finite, otherwise the division event will not be defined. Note that in general  $\vec{X}(T_{k+1}) \neq 2\vec{X}(T_k)$  and  $\Delta T_{k+1} \neq \Delta T_k$ , however we will prove in the next section that these conditions can be asymptotically approached.

Let us now use the hypothesis that the matrix  $M$  is invertible, so from Eqs. (2.4) we get:

$$(2.6) \quad \frac{dC}{dt} = \vec{\alpha} \cdot M^{-1} \frac{d\vec{X}}{dt},$$

hence the quantity  $Q(t) = C(t) - \vec{\alpha} \cdot M^{-1} \vec{X}(t)$ , is a first integral, i.e. a quantity constant during each division cycle (the proof is straightforward,  $dQ/dt = 0$  follows from Eq. (2.6)). Evaluating  $Q(t)$  at the beginning and at the end of the  $(k+1)$ -th



division we obtain

$$(2.7) \quad C(T_k) - \vec{\alpha} \cdot M^{-1} \vec{X}(T_k) = C(T_{k+1}) - \vec{\alpha} \cdot M^{-1} \vec{X}(T_{k+1}),$$

recalling that  $C$  takes the initial value  $\theta/2$  and the final value  $\theta$  and using the definition of  $\vec{X}_k$  (see Eq. (2.5)) we finally get:

$$(2.8) \quad \frac{\theta}{2} = \vec{\alpha} \cdot M^{-1} (2\vec{X}_{k+1} - \vec{X}_k).$$

Let us observe that this relation is meaningful only if the container size increases, otherwise the division event is not even defined and thus  $T_{k+1}$  is not formally defined.

On the other hand Eq. (2.4) can be explicitly solved during the  $(k+1)$ -th division cycle to give:

$$(2.9) \quad \vec{X}(t) = \vec{X}(t, T_k, \vec{X}_k) = e^{M(t-T_k)} \vec{X}_k,$$

hence we get the following relation between the amount of genetic material between two successive divisions:

$$(2.10) \quad \vec{X}_{k+1} = \frac{\vec{X}(T_{k+1})}{2} = \frac{1}{2} e^{M\Delta T_k} \vec{X}_k,$$

once again provided  $T_{k+1}$  exists.

We postpone the analysis of this model to Section 3, where we will be able to introduce a general framework where the SRM and the IRM can be put and solved.

**Remark 2.1.** *Let us comment on a simplification which has been previously used for the SRM model, namely the assumption that the surface is proportional to a power of the volume. This is certainly the case for a spherical micelle (with exponent 2/3), but in the case of a vesicle it holds (with exponent 1) only in the limit of a very large size and very thin thickness.*

*It can be shown that the finite size effects can be taken into account without modifying our results: synchronization is still obtained. In fact assuming a generic relation,  $S = f(C)$ , between the surface and the volume, and thus the container size, for some positive increasing function  $f$ , then Eqs. (2.3) have to be modified into:*

$$(2.11) \quad \frac{d\vec{X}}{dt} = \frac{f(C)}{C} M \vec{X} \quad \text{and} \quad \frac{dC}{dt} = \frac{f(C)}{C} \vec{\alpha} \cdot \vec{X}.$$

*But then we can observe that the function given by Eq. (2.6) is still a first integral and thus the same analysis follows. Another explanation of this result is that we can “rescale” the time<sup>3</sup> by the positive function  $C/f(C)$  and thus identifying Eqs. (2.11) and Eq. (2.6).*

**2.3. Internal Reaction Models of protocells.** In this section we will introduce and analyze a new family of protocell models inspired by the RNA-cell (Oberholzer et al., 1995; Szostak et al., 2001). We thus assume the protocell to be a vesicle made of lipidic material delimiting an inner water pool, where the relevant chemical reactions do occur. We also made the simplified assumption that precursor molecules, of both lipid precursor and genetic replicators, lying in the environment can enter and diffuse through the membrane fast enough to consider this step instantaneous.

---

<sup>3</sup>More precisely let us introduce a new non-linear time  $\tau = \int^t C^{-1}(s) f(C(s)) ds$  and let us denote the quantities  $C$  and  $\vec{X}$  using this new time, respectively by  $c(\tau)$  and  $\vec{x}(\tau)$ , then Eq. (2.11) is formally equivalent to Eq. (2.3).

We thus have a network of chemical reactions involving the reproduction of genetic material and the production of membrane molecules, running in the inner water pool. Let us consider, however, that the protocell increases its size – the membrane grows because the produced lipid molecules are supposed to be instantaneously incorporated into the membrane – hence we have to account for the volume variations into the “typical” kinetic equation used. We also assume the following *division hypothesis*: once the protocell membrane has doubled its size, with respect to the initial one, then protocell division occurs, guided by physical instabilities at the membrane level. At this point two identical protocells are obtained as offspring under the requirement of conservation of membrane and genetic molecules.

Let us start with the simplest case, assuming the protocell to be endowed with only one kind of genetic molecule, whose concentration at time  $t$  will be denoted by  $[X](t)$ , i.e. the ratio of the amount, or mass of  $X$ -molecules at time  $t$ ,  $X(t)$ , divided by the protocell volume,  $V(t)$ , enclosed by the membrane. We here hypothesize the membrane thickness to be very small in such a way that we can well approximate the internal volume with the total volume of the protocell, hence  $[X](t) = X(t)/V(t)$ . We also assume that the membrane surface  $S$  is proportional to the quantity of membrane molecules  $C$  and that the volume depends <sup>4</sup> on  $S$ .

Under the hypotheses of a linear growth for the reproduction of the genetic material and also that the membrane size growth is linear in the amount of  $X$ -molecules, we get the following system <sup>5</sup>:

$$(2.12) \quad \begin{cases} \frac{dC}{dt} &= \alpha[X]V \\ \frac{d[X]}{dt} &= \eta[X] - \frac{[X]}{V} \frac{dV}{dt}, \end{cases}$$

where  $\eta$  and  $\alpha$  are two positive coefficients taking into account respectively, the rate of self-replicating of  $X$  molecules and the rate of container growth, which also takes into account the amount of precursors that we assume to be constant. The last term in the second equation is due to the varying volume and we stress once again that we neglect the membrane thickness in computing the protocell volume.

Let us denote  $[X]_k$  the concentration of  $X$  at the beginning of the  $(k+1)$ -th division and once again  $\Delta T_k = T_{k+1} - T_k$  the duration of the  $(k+1)$ -th protocell cycle. The aim is to study the behaviour of  $[X]_k$  and  $\Delta T_k$  as a function of the division number and show, if any, the existence of some synchronization mechanism.

Let us start by solving the first equation (2.12). Because  $[X]$  is never zero, otherwise everything will stop, the solution describing the concentration of  $[X]$  during the  $k$ -th protocell cycle, is given by:

$$(2.13) \quad [X](t) = [X]_k \frac{V_{in}}{V(t)} e^{\eta(t-T_k)},$$

where  $V_{in} = V(T_k)$  is the initial volume and  $[X]_k$  the concentration of  $X$ -molecules at the beginning of the cycle, while  $V(t)$  is the protocell volume at time  $t$ . Because we ignore the time evolution of the volume, and we don't want to impose it, we have to use the remaining equation of (2.12), that can be easily rewritten using the

---

<sup>4</sup>If the protocell were be a sphere then  $V = (6\sqrt{\pi})^{-1} S^{3/2} \propto C^{3/2}$ .

<sup>5</sup>Let us observe that assuming the membrane thickness to be constant, one can express the surface of the protocell as a function of the container  $C$ ,  $S = C\delta$ , and thus the first equation in (2.12) can be rewritten in terms of geometric quantities as:  $\frac{dS}{dt} = \alpha'[X]V$ .

second one, as:

$$(2.14) \quad \frac{dC}{dt} = \frac{\alpha}{\eta} \frac{d}{dt} ([X](t)V(t)) .$$

Which implies the existence of a first integral:  $R(t) = C(t) - \frac{\alpha}{\eta}[X](t)V(t)$ . Hence equating  $R(t)$  at the beginning and the end of the  $(k+1)$ -th division we get:

$$(2.15) \quad C(T_{k+1}) - C(T_k) = \frac{\alpha}{\eta} ([X](T_{k+1})V(T_{k+1}) - [X](T_k)V(T_k)) .$$

The hypothesis that volume changes are dictated only by the surface variations, is based on the assumption of turgid vesicle, which in turn implies a relatively slow membrane growth, in such a way the volume can adjust to equilibrate the external and internal pressure. The explicit time variation of the volume, is thus limited to fluctuation around this equilibrium shape, that we neglect assuming it to be small.

We are now interested in determining the concentration of the  $X$ -molecule at the beginning of each division cycle as a function of the involved quantities. Recalling that, by the halving hypothesis, each offspring will be endowed with half the number of molecules produced in the previous cycle, and using moreover the fact that the concentration is obtained dividing this number by the volume at the beginning of the cycle, we obtain:

$$(2.16) \quad [X]_k = \frac{X(T_k)}{2} \frac{1}{V_{in}} = \frac{[X](T_k)V_{fin}}{2} \frac{1}{V_{in}} ,$$

where  $V_{fin}$  is the volume enclosed by the membrane surface just before the division. Finally using (2.15) and calling  $\sigma$  the threshold on the container size at which the protocell division occurs, we get:

$$\frac{\sigma}{2} = \frac{\alpha V_{in}}{\eta} (2[X]_{k+1} - [X]_k) .$$

This relation can be rewritten as:

$$[X]_{k+1} = \frac{[X]_k}{2} + \frac{\sigma\eta}{4\alpha V_{in}} ,$$

and then iterating back in time we get:

$$[X]_{k+1} = \frac{\sigma\eta}{4\alpha V_{in}} \sum_{j=0}^k \frac{1}{2^j} + \frac{[X]_0}{2^{k+1}} ,$$

which in the limit of infinitely many divisions converges to the asymptotic value:

$$[X]_{\infty} = \frac{\sigma\eta}{2\alpha V_{in}} ,$$

meaning that synchronization is obtained with a constant period given by:

$$\Delta T_{\infty} = \frac{1}{\eta} \log 2 .$$

Let us now consider the more general case where  $N$  kinds of different genetic molecules are present in the same protocell. Let us denote by  $[\vec{X}](t)$ , the vector of their concentrations at time  $t$ , i.e.  $[\vec{X}](t) = ([X^1](t), \dots, [X^N](t))$ .

Let us again assume linear interactions between the GMMs and, moreover, that the container growth is also linearly dependent on the amount of genetic molecules. Thus, we get the system:

$$(2.17) \quad \begin{cases} \frac{dC}{dt} &= \vec{\alpha} \cdot [\vec{X}]V \\ \frac{d[\vec{X}]}{dt} &= M[\vec{X}] - \frac{[\vec{X}]}{V} \frac{dV}{dt}, \end{cases}$$

where  $V$  is the volume enclosed by the surface membrane<sup>6</sup>,  $\vec{\alpha}$  is a  $N$ -dimensional vector with non-negative entries such that  $\vec{\alpha} \cdot [\vec{X}] = \sum_{i=1}^n \alpha_i [X^i]$  represents the contribution to the membrane growth due to  $[X^1](t), \dots, [X^N](t)$  and the (constant and real) matrix element  $M_{ij}$  represents the contribution of the  $X^j$ -molecule to the growth of the  $X^i$ -molecule. Once again we introduced a term due to the volume variations.

Using the assumption that the matrix  $M$  is non-singular, we can determine the first integral:

$$R(t) = C(t) - V(t)\vec{\alpha} \cdot M^{-1}[\vec{X}](t);$$

once again the following relation between the concentrations at the beginning of each cycle can be found:

$$[\vec{X}]_k = \frac{[\vec{X}](T_k)V_{fin}}{2} \frac{1}{V_{in}},$$

where  $V_{in}$  and  $V_{fin}$  have been defined previously. Calling  $\sigma$  the surface division threshold, we get:

$$(2.18) \quad \frac{\sigma}{2} = V_{in}\vec{\alpha} \cdot M^{-1} \left( 2[\vec{X}]_{k+1} - [\vec{X}]_k \right).$$

Once again this relation is meaningful only if the membrane is increasing, in such a way the existence of  $T_{k+1}$  is ensured.

Observe that the second Eq. (2.17) can be explicitly solved to give:

$$(2.19) \quad [\vec{X}](t) = \frac{V_{in}}{V(t)} e^{M(t-T_k)} [\vec{X}]_k,$$

where the volume variation in time is unknown, and thus:

$$(2.20) \quad [\vec{X}]_{k+1} = \frac{[\vec{X}](T_{k+1})}{2} = \frac{V_{in}}{2V_{fin}} e^{M\Delta T_k} [\vec{X}]_k,$$

provided  $T_{k+1}$  do exist.

We thus get a set of equations, cfr. (2.18) and (2.20), governing the evolution of the IRM, formally equal to the ones for the SRM, i.e. (2.8) and (2.10). These systems will be analyzed in the following section.

### 3. RESULTS : SYNCHRONIZATION IN LINEAR SRM AND IRM

The aim of this section is to present our main results, namely to prove under suitable hypotheses, the emergence of the synchronization phenomenon for both SRM and IRM whose physical relevance will be discussed in the next section, nevertheless we anticipate here that all the information is contained in the chemical constants, i.e. of the matrix  $M$  and the vector  $\vec{\alpha}$ , and our method allows us to extract it *a priori*, avoiding any numerical simulation. The first step is to rewrite

---

<sup>6</sup>Once again we assume that the membrane thickness can be neglected and assuming it constant we can rewrite the first relation in term of the membrane surface  $S$ .

both systems in a compact unified form. Roughly speaking this can be done because of the assumption of linear rate equations both for the genetic material and container growth. Formally let us introduce the following auxiliary notation:

	SRM	IRM
$\vec{\Xi}(t)$	$\vec{X}(t)$	$[\vec{X}](t)$
$\vec{\Xi}_k$	$\vec{X}_k$	$[\vec{X}]_k$
$a$	$1/2$	$V_{in}/(2V_{fin})$
$a'$	$1$	$V_{in}/V(t)$
$b$	$\theta/2$	$\sigma/(2V_{in})$

In fact (2.8) and (2.10) and (2.18) and (2.20) can be rewritten as:

$$(3.1) \quad \vec{\Xi}_{k+1} = ae^{M(t-T_k)}\vec{\Xi}_k \quad \text{and} \quad b = \vec{\alpha} \cdot M^{-1} \left( 2\vec{\Xi}_{k+1} - \vec{\Xi}_k \right).$$

Let us observe that the equations (2.9) and (2.19) governing the dynamics during each division cycle can also be cast in a compact form:

$$(3.2) \quad \vec{\Xi}(t) = 2a'e^{M(t-T_k)}\vec{\Xi}_k,$$

The aim of this section is to prove that synchronization holds provided  $M$  has an eigenvalue  $\lambda_1$  real, positive, with algebraic multiplicity 1 and possessing the largest real part, i.e.  $\lambda_1 > \Re \lambda_j$  for all the remaining eigenvalues  $\lambda_j$ . Moreover let  $\vec{v}_1$  be the eigenvector associated to  $\lambda_1$  then we assume that  $\vec{v}_1$  has positive entries <sup>7</sup>.

Thanks to the unified approach, we have just introduced, we can limit ourselves to consider only the SRM case and the IRM being completely analogous. The proof is thus obtained in two steps. First we prove in § 3.1 that under the previous assumptions there always exists a division event, namely for all  $k$  starting at time  $t = T_k$  there exists a positive lapse of time  $\Delta T_k = T_{k+1} - T_k$  such that  $C(T_k + \Delta T_k) = \theta$ . This is equivalent to requiring that the container grows,  $\dot{C} > 0$ . Second in § 3.2, we will prove that  $\vec{\Xi}_k$  eventually converges to  $\vec{w}$  where  $\vec{w}$  is proportional to  $\vec{v}_1$  and  $\vec{\alpha} \cdot \vec{w} = \lambda_1 b$ , and that the division intervals converge to a constant value  $\Delta T_k \rightarrow -\lambda_1^{-1} \log a$ , thus synchronization is achieved, in fact with our notations  $-\log a > 0$ . Thus the ultimate fate of both SRM and IRM is to synchronize the replication and division rates, moreover we can predict respectively the asymptotic amount of genetic material for SRM (and concentration of GMMs for IRM) in terms of the problem data.

Let us first give a proof in a simplified case which nonetheless contains all the main ideas and then leave complete proof to section 3.3.

**3.1. Analysis of the division time.** Let us assume  $M$  possesses  $N$  distinct eigenvalues,  $\lambda_1, \dots, \lambda_N$ . As previously stated we hypothesize that  $\lambda_1$  is real, positive, with algebraic multiplicity 1 and with the largest real part, i.e.  $\lambda_1 > \Re \lambda_j$  for all  $j = 2, \dots, N$ . Moreover we assume that its associated eigenvector  $\vec{v}_1$  has positive entries (see footnote 7).

By standard results of linear algebra the  $N$  eigenvectors of  $M$  define a basis of the whole space on which we decompose the vectors:

$$(3.3) \quad \vec{\Xi}_k = \xi_k^{(1)} \vec{v}_1 + \dots + \xi_k^{(N)} \vec{v}_N \quad \text{and} \quad \vec{\alpha} = \alpha^{(1)} \vec{v}_1 + \dots + \alpha^{(N)} \vec{v}_N,$$

<sup>7</sup>Because eigenvectors are defined up to multiplicative factors, by positive entries we mean that all entries have the same sign and thus we can multiply them to put them all positive.

then the solution of (3.2) can be rewritten as:

$$(3.4) \quad \vec{\Xi}(t) = 2a' \left( e^{\lambda_1(t-T_k)} \xi_k^{(1)} \vec{v}_1 + \dots + e^{\lambda_N(t-T_k)} \xi_k^{(N)} \vec{v}_N \right),$$

and we can easily observe that if  $t - T_k$  is large enough the first term involving the eigenvalue with largest real part will dominate. Hence the equations for the growth rate of the container can be rewritten as follows:

$$(3.5) \quad \dot{C} \sim 2a' e^{\lambda_1(t-T_k)} \xi_k^{(1)} \alpha^{(1)}.$$

Because  $\vec{\alpha}$  has positive entries we can assume generically that  $\alpha^{(1)} > 0$ , thus if  $t - T_k$  is large enough we get the growth of the container, i.e.

$$(3.6) \quad \dot{C} > 0,$$

which gives the desired result provided also  $\xi_k^{(1)} > 0$  for all  $k$ .

We are now able to prove by induction that the division event always exists. For the first division let us assume that the initial vector  $\vec{\Xi}_0$  has positive projection on  $\vec{v}_1$ , namely  $\xi_0^{(1)} > 0$ , then by (3.6) the container will grow and reach the division threshold at some time  $T_1 = T_0 + \Delta T_0$ , during this time  $\vec{\Xi}(t)$  always has a positive projection on  $\vec{v}_1$ , in fact  $\vec{\Xi}(t) \cdot \vec{v}_1 = 2a' e^{\lambda_1(t-T_0)} \xi_0^{(1)} > 0$ . Hence using the halving hypothesis the second cell cycle will start with a vector of initial conditions  $\vec{\Xi}_1$  still with a positive projection on  $\vec{v}_1$ , actually:  $\xi_1^{(1)} = a' e^{\lambda_1(t-T_0)} \xi_0^{(1)} > 0$ .

We now assume that the protocell undergoes  $k$  divisions as previously described, and we will prove that a further division will occur. So by hypothesis we start the  $k+1$  cycle with a vector of initial conditions  $\vec{\Xi}_{k+1}$  with a positive projection on  $\vec{v}_1$ ,  $\xi_{k+1}^{(1)} > 0$ , then from (3.6) we get that the protocell container is increasing and thus at some future time  $T_{k+1} = T_k + \Delta T_k$  it will reach the division threshold, the next cycle will start with a vector of initial conditions  $\vec{\Xi}_{k+1}$  with a positive projection on  $\vec{v}_1$ ,  $\xi_{k+1}^{(1)} = a' e^{\lambda_1 \Delta T_k} \xi_k^{(1)} > 0$ . Let us also observe that the division intervals are strictly positive,  $\Delta T_k > 0$  for all  $k$ . Otherwise, an infinite amount (or concentration) of genetic material will be required, against any physical reasonability.

This concludes thus the proof of the first part: starting with an initial amount (or concentration) of GMMs with a positive projection of the first eigenvector, the division mechanism never stops.

Let us now prove the claim about the asymptotic value of  $\vec{\Xi}_k$ .

**3.2. Asymptotic behaviour.** The aim of this section is to prove that  $\vec{\Xi}_k$  eventually converges to a well defined vector:  $\vec{w}$  where  $\vec{w} \in \text{span}(\vec{v}_1)$  and  $\vec{\alpha} \cdot \vec{w} = \lambda_1 b$ , and that the division intervals converge to a constant value  $\Delta T_k \rightarrow -\lambda_1^{-1} \log a$ , providing thus synchronization.

The second equation of (3.1) can be rewritten as:

$$(3.7) \quad \vec{\alpha} \cdot M^{-1} \left( 2\vec{\Xi}_{k+1} - \vec{\Xi}_k \right) = b = \vec{\alpha} \cdot \frac{\vec{w}}{\lambda_1} = \vec{\alpha} \cdot M^{-1} \vec{w},$$

thus introducing  $\vec{\delta}_k = \vec{\Xi}_k - \vec{w}$ , this last relation gives:

$$(3.8) \quad \vec{\alpha} \cdot M^{-1} \left( 2\vec{\delta}_{k+1} - \vec{\delta}_k \right) = 0,$$

namely

$$(3.9) \quad \vec{\alpha} \cdot M^{-1} \vec{\delta}_{k+1} = \frac{1}{2} \vec{\alpha} \cdot M^{-1} \vec{\delta}_k = \frac{1}{2^{k+1}} \vec{\alpha} \cdot M^{-1} \vec{\delta}_0,$$

so, in the limit of infinitely many divisions we conclude that:

$$(3.10) \quad \vec{\alpha} \cdot M^{-1} \vec{\delta}_{k+1} \rightarrow 0.$$

Once again assuming generically that  $\vec{\alpha}$  has a positive projection on  $\vec{v}_1$  (and thus also on  $\vec{w}$ ) and because  $M$  is invertible we conclude that:

$$(3.11) \quad \vec{\delta}_{k+1} = \vec{\Xi}_{k+1} - \vec{w} \rightarrow 0.$$

To prove that  $\Delta T_k \rightarrow -\frac{\log a}{\lambda_1}$  let us consider the first equation of (3.1) and rewrite it using the vector  $\vec{\delta}_k$ :

$$(3.12) \quad \vec{\delta}_{k+1} + \vec{w} = ae^{M\Delta T_k} \vec{\delta}_k + ae^{\lambda_1 \Delta T_k} \vec{w},$$

or, reordering the involved terms:

$$(3.13) \quad \vec{\delta}_{k+1} - ae^{M\Delta T_k} \vec{\delta}_k = (ae^{\lambda_1 \Delta T_k} - 1) \vec{w},$$

and in the limit of infinitely many divisions, recalling that  $\vec{\delta}_k \rightarrow 0$  we obtain:

$$(3.14) \quad (ae^{\lambda_1 \Delta T_k} - 1) \vec{w} \rightarrow 0,$$

hence  $\Delta T_k \rightarrow -\frac{\log a}{\lambda_1}$ .

**3.3. General case.** Let us now rapidly sketch the main changes one has to consider to deal with the general case where  $M$  has only  $N' < N$  eigenvalues, but still assuming that  $M$  has an eigenvalue  $\lambda_1$  real, positive, with algebraic multiplicity 1 and with the largest real part, i.e.  $\lambda_1 > \Re \lambda_j$  for all the remaining eigenvalues  $\lambda_j$ , and moreover its corresponding eigenvector has positive entries (see footnote 7).

The  $N'$  eigenvectors cannot define a basis of the whole space, but standard linear algebra results ensure that the set of  $N'$  vectors can be completed to give a basis using the Jordan vectors. Thus the decompositions (3.3) must be replaced with:

$$(3.15) \quad \vec{\Xi}_k = \xi_k^{(1)} \vec{v}_1 + \vec{\Xi}_k^\perp \quad \text{and} \quad \vec{\alpha} = \alpha^{(1)} \vec{v}_1 + \vec{\alpha}^\perp,$$

where  $\vec{\Xi}_k^\perp$  and  $\vec{\alpha}^\perp$  belong to the invariant subspace orthogonal to  $\vec{v}_1$ . But then using the hypothesis that  $\lambda_1$  has the largest real part and that the Jordan decomposition gives rise to invariant subspaces it is easy to prove the analogous of (3.6): if  $t - T_k$  is large enough the dynamics is still governed by  $\lambda_1$  and thus the container size increases provided  $\xi_k^{(1)}$  and  $\alpha^{(1)}$  are positive. As the proof continues similarly we therefore omit it.

In this section we have thus proved the following result:

**Theorem 3.1.** *Let  $\lambda_1$  be the eigenvalue of  $M$  with largest real part. If  $\lambda_1$  is positive, with algebraic multiplicity 1 and its associated eigenvector  $\vec{v}_1$  can be chosen with positive components. Then the SRM model (2.3) exhibits synchronization, provided  $\vec{\alpha} \cdot \vec{v}_1 > 0$  and  $\vec{X}_0 \cdot \vec{v}_1 > 0$ .*

*Moreover the asymptotic state is described by:*

$$SRM: \vec{X}_\infty = \|\vec{X}_\infty\| \vec{v}_1 : \|\vec{X}_\infty\| = \frac{\lambda_1 \theta}{2\vec{\alpha} \cdot \vec{v}_1} \quad \text{and} \quad \Delta T_\infty = \frac{\log 2}{\lambda_1} = \frac{\theta \log 2}{\vec{\alpha} \cdot \vec{X}_\infty}.$$

Let us observe that thanks to our unified approach we also have the following result:



**Theorem 3.2.** *Let  $\lambda_1$  be the eigenvalue of  $M$  with largest real part. If  $\lambda_1$  is positive, with algebraic multiplicity 1 and its associated eigenvector  $\vec{v}_1$  can be chosen with positive components. Then the IRM (2.12) do exhibit synchronization under the assumptions:  $\vec{\alpha} \cdot \vec{v}_1 > 0$  and  $[\vec{X}]_0 \cdot \vec{v}_1 \neq 0$ .*

*Moreover the asymptotic state is characterized by:*

$$\begin{aligned} \text{IRM: } [\vec{X}]_\infty = ||[\vec{X}]_\infty|| \vec{v}_1 : ||[\vec{X}]_\infty|| &= \frac{\lambda_1 \sigma}{2V_{in} \vec{\alpha} \cdot \vec{v}_1} \text{ and} \\ \Delta T_\infty &= \frac{1}{\lambda_1} \log \frac{2V_{fin}}{V_{in}} = \frac{\sigma}{2V_{in} \vec{\alpha} \cdot [\vec{X}]_\infty} \log \frac{2V_{fin}}{V_{in}}, \end{aligned}$$

where  $V_{in}$  respectively  $V_{fin}$  denote the protocell volume at the beginning and at the end of the division cycle.

The physical interpretation of the above results will be provided in § 4, here we limit ourselves to emphasize that from the knowledge of the chemical constants, i.e. the matrix  $M$  and the vector  $\vec{\alpha}$ , we can conclude if synchronization will arise or not. Moreover we are able to define the asymptotic division time and the amount of each genetic memory molecule. We'd also like to emphasize here that our analysis applies to models involving chemical reaction where replicators do not interact. More interesting cases, that can be described by our models, are represented by replicators that positively catalyze each other's synthesis but also inhibiting molecules are allowed provided their "influence" on the chemical network is small enough.

**3.4. Linear GARD – like models.** In this last part of the present section, we briefly introduce and discuss a class of protocells' models similar to the GARD – model (Sgré et al., 1998), where there is no longer a distinction between genetic material and container, namely the lipids themselves that form the protocell act also as information carriers: their relative amount determine the compositional information brought by the protocell (see Fig. 2 for a schematic representation of the model).

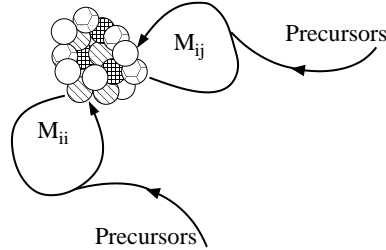


FIGURE 2. *A schematic representation of the Linear GARD-like model.*

We thus assume that lipid precursors are freely available (without any limitation) in the environment, and that they react with the lipids at the surface of the protocell to produce new lipids that are immediately incorporated into the protocell itself, which thus grows in size. We assume that once the size has reached some threshold, say twice the initial one, then the protocell splits into two identical offspring halving the mother lipid content (i.e. the compositional information). This model can thus

be casted in the SRM class. The amount of the  $i$ -th lipid,  $X^{(i)}$ , evolves in time according to (see Eq. (2.3)):

$$(3.16) \quad \frac{dX^{(i)}}{dt} = \sum_{j=1}^N M_{ij} X^{(j)} C^{\beta-1},$$

except that now the container size is given by:

$$(3.17) \quad C(t) = X^{(1)}(t) + \dots + X^{(N)}(t) = \vec{\alpha} \cdot \vec{X}(t),$$

where we introduced  $\vec{\alpha} = (1, \dots, 1)$ , i.e. the total amount of lipid molecules. We call these models *Linear GARD-like* because the chemical reactions involved in our model are linear in contrast with the GARD model where quadratic reactions have also been considered.

Because the container size is always positive we can “rescale” the time so as to include the factor  $C^{\beta-1}$  into the new time, hence the behaviour can be described by:

$$(3.18) \quad \frac{d\vec{X}}{dt} = M\vec{X} \quad \text{and} \quad C(t) = \vec{\alpha} \cdot \vec{X}(t).$$

The division and halving hypotheses imply that the amount of lipids between two successive cycles are related by the following law:

$$(3.19) \quad \theta = C(T_{k+1}) = \vec{\alpha} \cdot \vec{X}(T_{k+1}) = 2\vec{\alpha} \cdot \vec{X}_{k+1},$$

but

$$(3.20) \quad \frac{\theta}{2} = C(T_k) = \vec{\alpha} \cdot \vec{X}_k,$$

thus

$$(3.21) \quad \vec{\alpha} \cdot \vec{X}_{k+1} = \vec{\alpha} \cdot \vec{X}_k.$$

Calling  $\lambda_1$  the real, positive eigenvalue of  $M$  with largest real part and  $\vec{v}_1$  its associated eigenvector, with positive entries, we can conclude following the previous analysis for the SRM models, that the long term dynamics is completely determined by this eigenvalue and that the asymptotic amount of lipids will have a positive projection, only on the direction of  $\vec{v}_1$ , its absolute value being determined by (3.21). Namely:

$$(3.22) \quad \vec{X}_k \rightarrow \vec{X}_\infty = \|\vec{X}_\infty\| \vec{v}_1 \quad \text{such that} \quad \|\vec{X}_\infty\| = \frac{\theta}{2\vec{\alpha} \cdot \vec{v}_1},$$

and

$$(3.23) \quad \Delta T_k \rightarrow \Delta T_\infty = \frac{\log 2}{\lambda_1}.$$

#### 4. DISCUSSION

Let us now analyze the physical conditions ensuring that the matrix  $M$  has a single eigenvalue with largest real part (ELRP for short) and a corresponding positive eigenvector.

We first discuss the important case where all the matrix elements are non negative, i.e  $M_{ij} \geq 0$ , for all  $i, j = 1 \dots, N$ . This implies that there is *no negative interference* between different replicators  $i$  and  $j$ , the only possible alternatives being that either  $i$  favors (e.g. catalyzes) the formation of  $j$  or that it does not

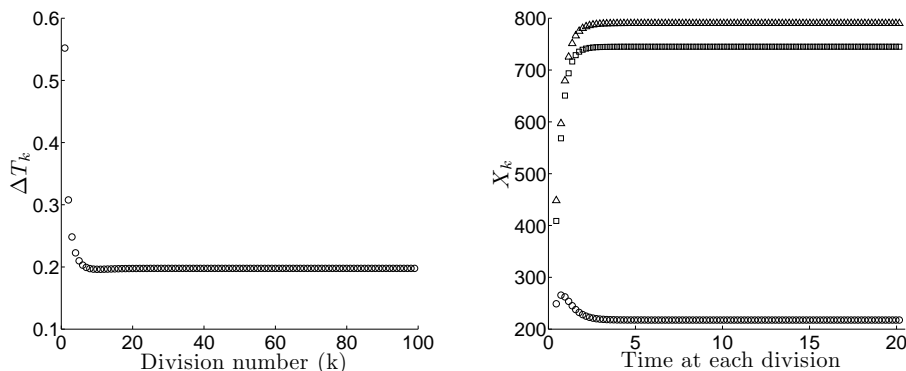


FIGURE 3. Numerical simulations of the SRM system described by Eq. (2.3) with a positive matrix  $M$  given by (4.1), whose ELRP is  $\lambda_1 = 3.5054$ . Left panel the division time,  $\Delta T_k$ , is plotted as a function of the generation number. On the right panel we report the amount of genetic material  $X^{(1)}$ ,  $X^{(2)}$ ,  $X^{(3)}$  at the beginning of each division cycle in function of the division number. Symbols refer to :  $X^{(1)}$  ( $\triangle$ ),  $X^{(2)}$  ( $\square$ ) and  $X^{(3)}$  ( $\circ$ ).

influence it in any way. Moreover, we must also require that at least one of the entries  $M_{ij}$  does not vanish, since otherwise there would be no replication at all. We can therefore apply the Perron–Frobenius theorem (Lütkepohl, 1996; Minc, 1988), which states that under the previous assumptions then the eigenvalue with the largest module is real, positive and unique, and that there is a non-negative eigenvector belonging to that eigenvalue.

Let us emphasize here that our method enables us to also predict which replicator molecules will go extinct, in fact once the main eigenvector has zero entries, the corresponding  $X^{(i)}$ –molecule will be absent from the asymptotic state.

We report in Fig 3 a numerical simulation for a positive matrix, thus casting in the Perron–Frobenius setting, in the case of a SRM model:

$$(4.1) \quad M = \begin{pmatrix} 1.6294 & 1.8268 & 0.5570 \\ 1.8116 & 1.2647 & 1.0938 \\ 0.2540 & 0.1951 & 1.9150 \end{pmatrix},$$

whose eigenvalues are  $\lambda_1 = 3.5054$ ,  $\lambda_2 = 1.6797$  and  $\lambda_3 = -0.3759$ , and corresponding eigenvectors  $\vec{v}_1 = (0.7134, 0.6727, 0.1965)$ ,  $\vec{v}_2 = (0.5420, 0.2587, -0.7996)$  and  $\vec{v}_3 = (0.6752, -0.7375, -0.0121)$ . According to Theorem 3.1 the asymptotic division time is  $\Delta T_\infty = \log 2 / \lambda_1$  in excellent accord with the numerical value  $\sim 0.1977$ , and the asymptotic amount of genetic material  $\vec{X}_\infty = \lambda_1 \theta / (2\vec{\alpha} \cdot \vec{v}_1)$  is again in excellent accord with the numerical value  $(790.0772, 745.0027, 217.6201)$ . The simulation has been performed using the values  $\theta = 1000$ ,  $\vec{\alpha} = (1, 1, 1)$  and  $\vec{X}_0 = (191.4334, 97.0751, 160.0561)$ . It can easily be verified that both have non zero projection on  $\vec{v}_1$ .

One can prove that synchronization arises under weaker conditions, in particular the matrix  $M$  can have negative entries, but then the *autocatalytic contributions*, i.e.  $M_{ii} > 0$ , must be *large enough* with respect to the possible inhibitory terms,  $M_{ij}$  with  $j \neq i$ . More precisely using the *Gershgorin circle theorem* (Golub and Van Loab, 1996) and the assumption that  $M$  is a strictly diagonally dominant

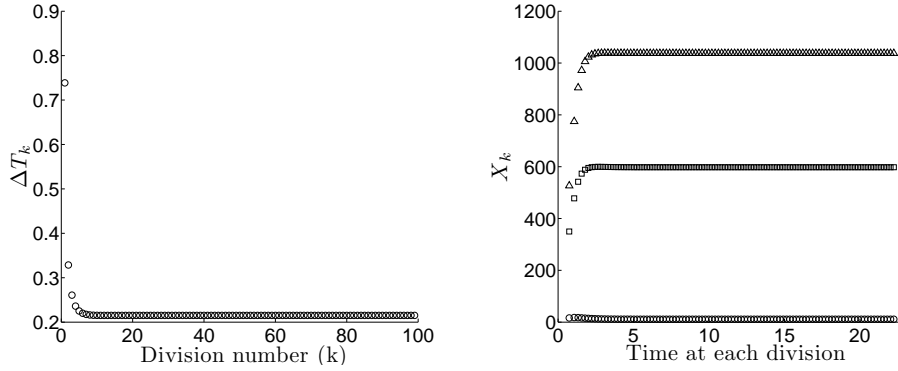


FIGURE 4. Numerical simulations of the SRM system described by Eq. (2.3) with a diagonal dominant matrix  $M$  given by (4.2), whose ELRP is  $\lambda_1 = 3.2209$ . Left panel the division time,  $\Delta T_k$ , is plotted as a function of the generation number. On the right panel we report the amount of genetic material  $X^{(1)}, X^{(2)}, X^{(3)}$  at the beginning of each division cycle in function of the division number. Symbols refer to :  $X^{(1)}$  ( $\triangle$ ),  $X^{(2)}$  ( $\square$ ) and  $X^{(3)}$  ( $\circ$ ).

matrix, i.e. for all  $i$ ,  $|M_{ii}| > \sum_{j \neq i} |M_{ij}|$ , then once again one has synchronization provided the off diagonal terms are small enough. Of course the assumption on the positiveness of the component of the eigenvector is still needed.

We report in Fig 4 a numerical simulation for a strictly dominant matrix, where the eigenvector associated to the real and largest eigenvalue has positive entries, still for the SRM model:

$$(4.2) \quad M = \begin{pmatrix} 2.6294 & 1.02688 & -0.0055 \\ 1.1161 & 1.2647 & 1.0038 \\ 0.0254 & -0.0195 & 1.9150 \end{pmatrix},$$

whose eigenvalues are  $\lambda_1 = 3.2209$ ,  $\lambda_2 = 0.6985$  and  $\lambda_3 = 1.8897$  and corresponding eigenvectors  $\vec{v}_1 = (0.8665, 0.4992, 0.0094)$ ,  $\vec{v}_2 = (0.4693, -0.8827, -0.0239)$  and  $\vec{v}_3 = (-0.5011, 0.3652, 0.7846)$ . According to Theorem 3.1 the asymptotic division time is  $\Delta T_\infty = \log 2 / \lambda_1$  in excellent accord with the numerical value  $\sim 0.2152$ , and the asymptotic amount of genetic material  $\vec{X}_\infty = \lambda_1 \theta / (2\vec{\alpha} \cdot \vec{v}_1)$  is again in excellent accord with the numerical value  $(1014.8, 584.6, 11.0)$ . The simulation has been performed using the values  $\theta = 1000$ ,  $\vec{\alpha} = (1, 1, 1)$  and  $\vec{X}_0 = (162.9447, 181.1584, 25.3974)$ . It can easily be verified that both have non zero projection on  $\vec{v}_1$ .

We observed that for generic matrices, i.e. for  $M_{ij}$  that can have both positive and negative signs, with no particular relations between their values, synchronization is not always reached, even if there are positive eigenvalues. Let us thus consider the case where some entries of the real matrix  $M$  can be negative, while it still possesses  $N$  independent eigenvectors. In this case  $M$  can still be diagonalized and therefore the ELRP determines the long term behaviour of the system. However now in general: i) the eigenvector associated to the ELRP may have negative components and/or ii) the ELRP may be complex, so the previous equations loose their physical meaning.

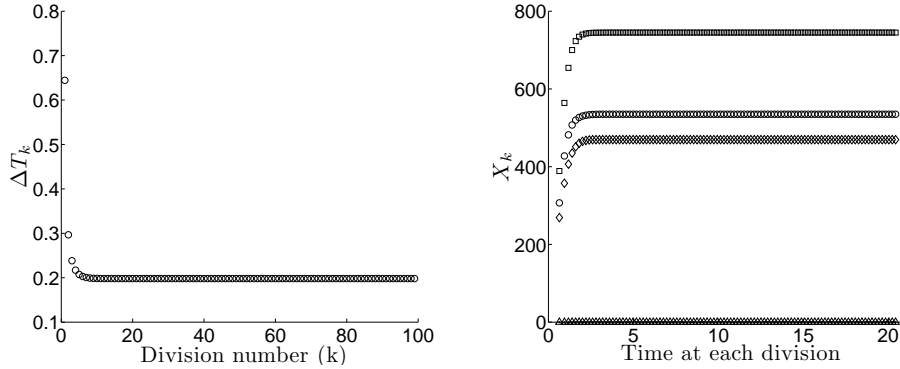


FIGURE 5. *Some chemicals are definitely extinguished.* An example of a  $5 \times 5$  matrix  $M$  with negative entries: it possesses a real, positive eigenvalue with the largest real part but the corresponding eigenvector has both positive and negative components. The replicators which survive are those which might have been predicted by inspection of the eigenvector associated to the ELRP. Symbols refer to :  $X^{(1)}$  ( $\bullet$ ),  $X^{(2)}$  ( $\circ$ ),  $X^{(3)}$  ( $\square$ ),  $X^{(4)}$  ( $\diamond$ ) and  $X^{(5)}$  ( $\triangle$ ).

Let us first consider the case of a real eigenvalue whose eigenvector has positive and negative components. A possible hypothesis could be to try to extend the theory to deal with these cases by assuming that, whenever one of  $X^{(i)}$ 's becomes negative, it has to be interpreted as being actually equal to zero (the non-physical negative value indicating some limitation of the model used which relies on an ODE system). The rationale is that if the amount of the  $X^{(i)}$ -molecule, starting from a positive value, “becomes negative”, it must have passed through the value 0: in this case there is no more replicator in the system, and it is justified to set its value equal to 0. The value of  $X^{(i)}$  may become positive again at a later time if it is produced by reactions involving other replicators.

Since the analytical theory is not applicable we performed some numerical analyses to get some new insight. Thus we first consider the case where some components get permanently extinguished. If we drop from the matrix  $M$  those components which the simulation shows go to extinction, we obtain a reduced matrix  $M'$ . If its ELRP is positive and its eigenvector non-negative then the previous analytical theory applies and correctly predicts asymptotic duplication time and quantities of replicators, as reported for instance in Fig. 5.

We also analyzed the case where some components can be recreated by the chemical network. While some simulations show synchronization, we have indeed also found some different behaviours, where the duplication time does not reach a constant value but seems to oscillate periodically in time, see for instance Fig. 6.

Briefly, one can conclude that the analytical method precisely describes the system behaviour when the ELRP is real and positive and its eigenvector non-negative. In different cases one has to resort to simulations, however the analytical theory may still help in understanding the system's behaviour (like in the case where the reduced matrix  $M'$  has the properties required to apply it).

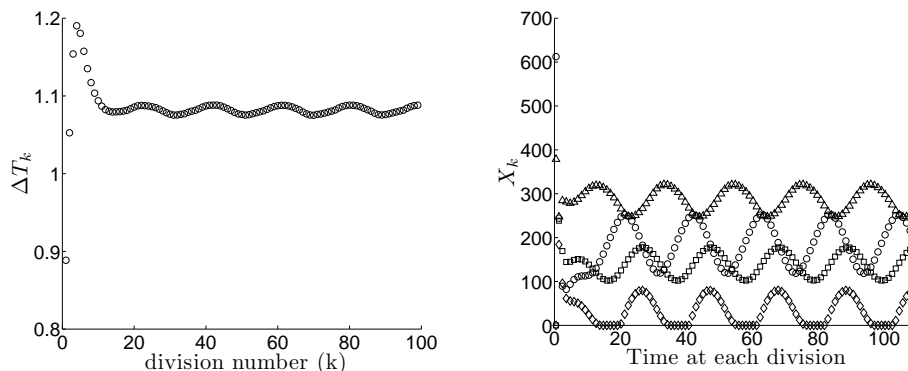


FIGURE 6. *Extinguished chemicals can be recreated by the reaction network.* An example of a  $4 \times 4$  matrix  $M$  with negative entries where synchronization is not achieved but the division time and the initial amount of genetic material regularly oscillate division after division. Right panel: the time evolution behaviour of the amounts of  $\bar{X}_k$  at the beginning of each division and Left panel the division time,  $\Delta T_k$ , in function of the generation number  $k$ . Symbols refer to :  $X^{(1)}$  ( $\triangle$ ),  $X^{(2)}$  ( $\square$ ),  $X^{(3)}$  ( $\circ$ ) and  $X^{(4)}$  ( $\diamond$ ).

So far we have considered the case where there is a single ELRP, or a pair of complex conjugate ELRPs. If there were more ELRPs then the above analysis has to be extended, as it should be done to deal with the case where the ELRP were a multiple eigenvalue.

**4.1. Conclusions.** In the present paper we address some relevant questions about the synchronization phenomenon for systems where the kinetic equations are linear, and for several analyzed architectures we provided sufficient conditions to ensure synchronization, which thus result in an emergent property of the model. Even if our analysis is able to cover several relevant cases, some further investigations are needed to understand some remaining cases. Moreover we are aware that non-linear terms may play a key role and thus further studies are surely necessary to give a comprehensive account of the behaviour of non-linear systems.

*Acknowledgment* Support from the EU FET-PACE project within the 6th Framework Programme under contract FP6-002035 (Programmable Artificial Cell Evolution) is gratefully acknowledged. Also Elena Lynch of the European Center for Living Technologies is kindly acknowledged for the final redaction of the manuscript.

## REFERENCES

- Bartel, D.P., 1999, *The RNA World*, Eds Gesteland, R.F., Cech, T.R. and Atkins, J.F., pp. 143 (Cold Spring Harbor Laboratory Press, New York, 1999).
- Bartel, D.P. and Unrau, P.J., 1999, Constructing an RNA world, *Trends Cell Biol.* **9**, 9–13.
- Carletti, T. and Fanelli, D., 2007, From chemical reactions to evolution: emergence of species, *Europhys. Letters* **77**, 18005p1–p6.

- Gánti, T., 1997, *J. Theor. Biol.* **187**, 583–593.
- Golub, G.H. and Van Loan, C.F., 1996, *Matrix computations*, 3d Ed, (Johns Hopkins University press, Baltimore USA 1996).
- Hager, H.J., Pollard, J.D. and Szostak, J.W., 1996, Ribozymes: aiming at RNA replication and protein synthesis, *Chem. Biol.* **3**, 717–725.
- Kaneko, K. and Yomo, T., 2002, On a kinetic Origin of Heredity: Minority Control in a Relicating System with Mutually Catalytic Molecules, *J. Theor. Biol.* **214**, 563–576., doi:10.1006/j.jtbi.2001.2481.
- Kaneko, K., 2006, *Life : An introduction to complex system biology*, (Springer, The Netherlands 2006)
- Lütkepohl, H., 1996, *Handbook of matrices*, (John Wiley, New York, 1996)
- Minc, H., 1988, *Nonnegative matrices*, (John Wiley, New York, 1988)
- Munteanu, A. et al., 2006, Generic Darwinian selection in protocell assemblies, SFI Working Paper **06-09-032** (2006)
- Munteanu, A. and Solé, R.V., 2006, Phenotypic diversity and chaos in a minimal cell model, *J. Theor. Biol.* **240**, 434–442., doi:10.1016/j.jtbi.2005.10.013
- Nelson, K.E., Levy, M., and Miller, S.L., 2000, Peptide nucleic acids rather than RNA may have been the first genetic molecule, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8, 3868–3871.
- Oberholzer, T. et al., 1995, Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell, *Biochemical and Biophysical Research Communications* **207**, 250–257.
- Rasmussen, S. et al., 2003, Bridging nonliving and living matter, *Artificial Life* **9**, 269–316.
- Rasmussen, S. et al., 2004, Proto-Organism kinetics: evolutionary dynamics of lipid aggregates with genes and metabolism, *Origins of Life and Evolution of the Biosphere* **34**, pp. 171–180.
- Segré, D. et al., 1998, Graded autocatalysis replication domain (GARD): kinetic analysis of self-replication in mutually catalytic sets, *Origins of Life and Evolution of the Biosphere* **28**, 501–514.
- Serra, R., Carletti, T., Poli, I., 2007, Synchronization Phenomena in Surface-Reaction Models of Protocells, *Artificial Life* **13**, 123–138.
- Serra, R., Carletti, T., Poli, I., 2007, Surface-reaction models of protocells, in *BIOMAT2006* Ed. Mondaini, R.P. and Dilão, R., (World Scientific Co. Pte. Ltd. 2007)
- Szostak, D., Bartel, P.B. and Luisi, P.L., 2001, Synthesizing life, *Nature* **409**, 387–390.
- Mansy, S.S. et al., 2008, Template-directed synthesis of a genetic polymer in a model protocell, *Nature Letters*, doi:10.1038/nature07018.

(Timoteo Carletti) DÉPARTEMENT DE MATHÉMATIQUE, FACULTÉS UNIVERSITAIRES NOTRE DAME DE LA PAIX, REMPART DE LA VIERGE 8, B 5000 NAMUR, BELGIUM AND DIPARTIMENTO DI STATISTICA, UNIVERSITÀ CA' FOSCARI, SAN GIOBBE - CANNAREGIO 873, 30121 VENEZIA, ITALY

(Roberto Serra) DIPARTIMENTO DI SCIENZE SOCIALI, COGNITIVE E QUANTITATIVE, UNIVERSITÀ DI MODENA E REGGIO EMILIA, VIA ALLEGRI 9, 42100 REGGIO EMILIA, ITALY AND DIPARTIMENTO DI STATISTICA, UNIVERSITÀ CA' FOSCARI, SAN GIOBBE - CANNAREGIO 873, 30121 VENEZIA, ITALY

(Irene Poli) DIPARTIMENTO DI STATISTICA, UNIVERSITÀ CA' FOSCARI, SAN GIOBBE - CANNAREGIO 873, 30121 VENEZIA, ITALY



(Marco Villani) DIPARTIMENTO DI SCIENZE SOCIALI, COGNITIVE E QUANTITATIVE, UNIVERSITÀ DI MODENA E REGGIO EMILIA, VIA ALLEGRI 9, 42100 REGGIO EMILIA, ITALY

(Alessandro Filisetti) EUROPEAN CENTER FOR LIVING TECHNOLOGY, CALLE DEL CLERO 2940, 30124 VENEZIA, ITALY

*E-mail address*, Timoteo Carletti: `timoteo.carletti@fundp.ac.be`

*E-mail address*, Roberto Serra: `rserra@unimore.it`

*E-mail address*, Irene Poli: `irenpoli@unive.it`

*E-mail address*, Marco Villani: `villani.marco@unimore.it`

*E-mail address*, Alessandro Filisetti: `alessandro.filisetti@ecltech.org`